IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

PREPROCESSING TECHNIQUES FOR MONEY LAUNDERING PROCESS

[1]Dr Pushpa Ravikumar, [2]Leela C P, [3]Chaithra I V, [4]Sangareddy B Kurtakoti
[1]Professor & Head, [2]M.Tech scholar, [3]Asst. Professor, [4]Asst. Professor
Computer Science and Engineering
Adichunchanagiri Institute of Technology, Chikkamagaluru, Karanataka, India

Abstract : Money laundering is the unlawful act of disguising the source of money gained unlawfully by transferring it via a convoluted series of financial transfers or business dealings. Overall, this procedure insinuates and obliquely returns the "clean" money to the launderer. The technique of altering the colour of money is known as money laundering. Additionally, it is a finance sector virus. The placement, layering, and integration are the three stages. Around the world, there are several ways to detect money laundering, at different stages. Here is our new method, which uses deep learning in its initial stage and is more accurate. This benefits a wide range of financial organisations and financial sectors, especially banks.

IndexTerms **- Money Laundering, Decision Tree, Random Forest, Artificial Neural Network**

I. INTRODUCTION

Moving unlawfully obtained money through a web of complex financial transactions to seem legitimate or "clean" is known as money laundering. The purpose of money laundering is to conceal the real source of the funds, making it difficult for law authorities to connect them to their illegal source. Money laundering is the process of transferring monies that have been obtained unlawfully via a series of complex financial transactions in an effort to make them seem genuine or "clean." The purpose of money laundering is to conceal the real source of the funds, making it difficult for law authorities to connect them to their illegal source. Because of this, criminals may generate money from their illegal actions without drawing attention to themselves. But it is the responsibility of the financial sectors and institute to detect the illegal activity at the different stages of the process.

Through this mechanism, tax-avoidance income imitates actual income. Money travels through the whole global financial system before it reaches the account of the intended beneficiary. It may be challenging to trace the flow of money used for illegal activities, especially in the future. There are several methods to commit financial crimes using the monetary system of the economy [1][5][16].

Money may be brought into the financial system through a number of techniques, such as smurfing. In order to modify the money's form, transfers will be made between different bank accounts. This will aid in the formation of a complicated layer that will obscure the origin of the cash and make their monitoring difficult. Deep learning is used in the identification and halting of money laundering because it helps financial organizations see questionable behaviors and trends in massive amounts of transactional data [1].

II. NEED OF THE STUDY.

1.1 Money Laundering

In to the financial system Money can be introduced in a variety of methods, including smurfing. Additionally, the money will be transferred between various bank accounts to change its form. This will contribute to the development of a convoluted layer that will conceal the source of the funds and make their tracking challenging.

1.2 Machine Learning

It can assist financial institutions in identifying suspicious behaviours and patterns in high volumes of transactional data, machine learning is employed in the detection and stopping the laundering of money.

1.3 Deep Learning

Deep learning is used in the identification and halting of money laundering because it helps financial organizations see questionable behaviors and trends in massive amounts of transactional data

3.1 Literature survey

We review the literature in the ML field in this section, and we wrap up the description of various techniques in this section. The process of converting unclean money that looks to have originated from a legitimate source into clean money is known as money laundering.

Despite the mounting issues of financial scandals, money laundering, and terrorism funding, the financial industry is expanding all over the world. The inadequacies of the financial system have received a lot of attention. This evaluation looks at previous research, knowledge progress, and real-world applications in the fight against money laundering and terrorism support. Additionally, future study recommendations are offered, as well as information on gaps in the preventative measures put in place by governments to combat anti-terrorism funding (ATF) and anti-money laundering (AML) [1].

Money laundering is the process of converting illegal monies that appear to have originated from a trustworthy source into legal cash. To put it another way, it introduces unlawfully obtained monies into the regular financial cycle or money circulation process while masquerading as legal tender. According to this study, the SARS has a largely negative impact, and corruption and the SARS are two concurrent factors impacting the money laundering issue [2].

Money laundering has been a global problem for decades and is a huge danger to the economy and society. Several studies have been undertaken on the use of ML for Anti-Money Laundering (AML) and Explainable Artificial Intelligence (XAI) approaches in general, but there has been no research on the use of DL techniques in conjunction with XAI. The purpose of this study is to examine the present state-of-the-art literature on DL and XAI for detecting suspect money laundering transactions and to recommend future research topics. This research work includes Convolutional Neural Networks and AutoEncoder[3].

A thorough assessment of the literature on money laundering was done, with special emphasis placed on the databases Pro-Quest, Scopus, and Science-Direct. Following a review of the literature, major study themes were identified. The topic of identifying money laundering was then thoroughly researched. The main strategies for this type of detection that employ ML and DL have been discovered [4].

The article looks at worldwide principles for fighting money laundering and other forms of corruption. Aside from their breadth, various provisions of the aforementioned international legal acts control the investigation of crimes involving money laundering, terrorist financing, and corruption. The International Financial Acts definition [5].

This paper's goal is to increase the application module's correctness. The paper is organized as, section I is with literature survey, II gives the research methodology , Machine Learning algorithms employed in III, IV with result and analysis, section V represents conclusion and VI is with the references.

III. RESEARCH METHODOLOGY

3.1 Collection Of Data

The research data was derived through a simulation of money-laundering activities in Middle Eastern banks using actual datasets. Utilising features and Information that is comparable to actual transfers from the original dataset, the data were simulated. The simulation's features and the procedures utilised in actual transactions had a lot in common. Both money laundering and non-money laundering were produced after consideration. A rather comprehensive simulator was created by attempting to model both money laundering and non-money laundering components of transactions. The simulation is built on the three methods used by financial organisations to launder money: placement, layering, and integration. A rule was established to represent cash-in transfers and another to represent transferred-out funds in the simulation of each procedure. An essential component of the advantage of a stimulated dataset is that it is adaptable and generates a dataset with various parameters.

3.1.1. Transaction Type

Either cash-in or transfer-out transactions fall under this category. Transaction type was codified as a category variable before being changed into dummy variables.

3.1.2. Amount Of Crime

Whether the money-laundering activities were carried out by the financial institution's CEO or a fellow employee depends on the severity of crime. A categorical variable that represented the level of criminality was coded before being transformed into dummy variables.

3.1.3. Quantity Of Money

The quantity of money, is a continuous random variable, indicates the actual monetary amount that was sent through the transaction.

3.1.4. Date

The transaction's date is just the day, month, and year. Python's date and time formats were used to further reformat the date by adding categorical variables to represent the days of the week and the months of the year.

3.1.5. Time

A continuous number, rounded to the nearest hour, represents the transaction's time.

3.1.6. Pofession of the account holder

The depending on this field the threshold is fixed. This field is grouped in to 2 groups and according to that the value is assigned in the preprocessing stage.

3.1.7. Old balance of the source

The old balance of the money sender is recorded before the transaction is performed.

3.1.8. New balance of the source

The new balance of the money sender is recorded after the transaction is performed.

3.1.9. Old balance of the destination

The old balance of the money receiver is recorded before the transaction is performed.

3.1.10. New balance of the source

The new balance of the money reciver is recorded after the transaction is performed.

3.1.11. Label

Whether or whether there was money laundering was the aim variable. The target variable was programmed to take the value 1 for a money laundering classification and 0 for a non-money laundering classification. Equation 1 displays the formula to represent the target variable.
Fraud = 1 and Legitimate = 0.

3.2 Statistical Tool And Performance Metrics

For the coding and analysis Python programming in a Jupyter notebook is employed. The chosen library was Scikit-learn. A well-known ML library for creating and analysing ML algorithms is called Scikit-learn. For the ANN analysis, the Keras opensource software library was utilised. The accuracy scores were used to assess the ML algorithm and ANN model. The confusion matrix was used to identify false positives or type 1 errors since this is a classification model. The most effective classifier was determined to be the algorithm with the highest predicted accuracy. The effectiveness of the top classification algorithm on the test set (or unseen data) was evaluated using a confusion matrix. The following performance criteria from a categorization report were also used to assess if the To determine if the model included the money-laundering category and the non- money-laundering categorization, precision, recall, and the F-1 score were also used. The trade- off between the false positive and the actual positive was plotted using the Receiver Operating Characteristic (ROC) evaluation tool.

3.3 Data Preprocessing

During the data preparation step, the dataset was checked for any unreported values. The crime level characteristic is absent in 38% of the observations. Because this is a categorical variable, the mode was used to impute the missing observations. The duplicate values in the dataset were all deleted. The individual and distinctive values of each characteristic were identified. When Not a Number (nan) values were discovered, they were replaced with zero. Among the characteristics removed from the dataset were source and destination IDs. The data could not be used to determine the location (i.e., latitude and longitude).
By creating a copy of the DataFrame, one of the copies is sliced into training (80%) and testing (20%).

1. The StandardScaler object is created and the training dataset is fitted.
2. By using scalar.transform the traning and testing features are transformed to fit the scaler.
3. Next, by using keras.Tokenizer to tokenize the textual columns and then, keras.pad_sequences is used to pad the column sequence to the same length.
4. Finally for the left of DataSet, the data cleaning is applied to clean out he null and unwanted columns.

This whole process is known as data pruning.

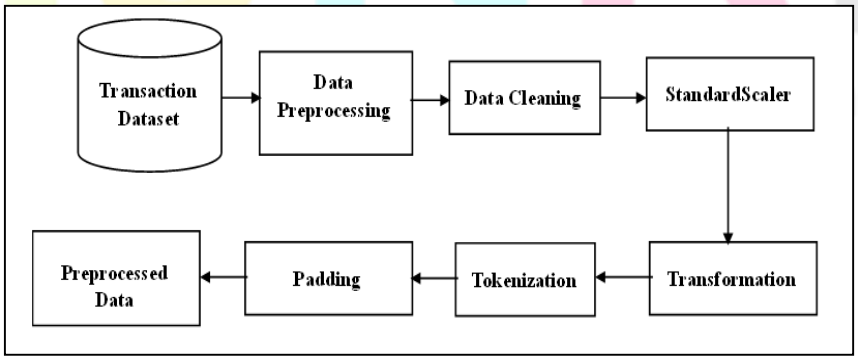


Fig. 1. Preprocessing System Architecture

In Fig 1,the transaction dataset is subjected to Data cleaning, the process of detecting and repairing flaws or inconsistencies in data, such as missing numbers, outliers, and duplicates. For data cleaning, many strategies such as imputation, elimination, and

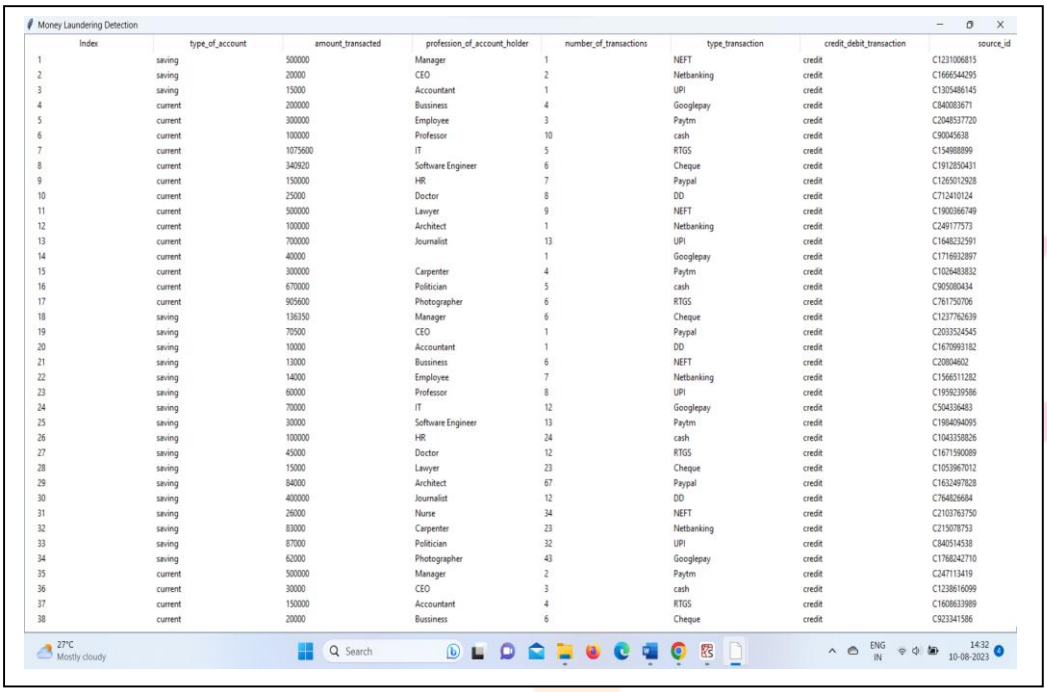
transformation can be utilised. Here, transaction dataset is freed from unnecessary values, then the duplicate data is removed, the textual data is converted to numerical values like type_of_account, profession of the account holder etc., searched for missing values if required whole tupple is removed finaly outliers are removed.

The mean is removed and each feature/variable is scaled to unit variance using StandardScaler. This procedure is carried out in a feature-by-feature manner. Because it includes estimating the empirical mean and standard deviation of each feature, StandardScaler can be impacted by outliers (assuming they exist in the dataset).

The transform() function is used to invoke a function on self, which results in a Series with transformed values and the same axis length as self. To be used for data transformation.

By using Tokenization technique the whole dataset each attribute is converted in to tokens. Then sequence. pad_sequences is a function for padding sequences. This function converts a list of num_samples sequences (integer lists) into a 2D Numpy array with the shape (num_samples, num_timesteps). num_timesteps is either the maxlen parameter if one is supplied, or the length of the longest sequence if none is supplied, as shown in the Fig 1. Finally the preprocessed data is generated.

IV. RESULTS AND ANALYSIS



Index	type_of_account	amount_transacted	profession_of_account_holder	number_of_transactions	type_transaction	credit_debit_transaction	source_id
1	saving	500000	Manager	1	NEFT	credit	C1231006815
2	saving	20000	CEO	2	Netbanking	credit	C1666544295
3	saving	15000	Accountant	1	UPI	credit	C1305486145
4	current	200000	Bussiness	4	Googlepay	credit	C840083671
5	current	300000	Employee	3	Paytm	credit	C2048537720
6	current	100000	Professor	10	cash	credit	C90045638
7	current	1075600	IT	5	RTGS	credit	C154988899
8	current	340920	Software Engineer	6	Cheque	credit	C1912850431
9	current	150000	HR	7	Paypal	credit	C1265012928
10	current	25000	Doctor	8	DD	credit	C712410124
11	current	500000	Lawyer	9	NEFT	credit	C1900366749
12	current	100000	Architect	1	Netbanking	credit	C249177573
13	current	700000	Journalist	13	UPI	credit	C1648232591
14	current	40000		1	Googlepay	credit	C1716932897
15	current	300000	Carpenter	4	Paytm	credit	C1026483832
16	current	670000	Politician	5	cash	credit	C905080434
17	current	905600	Photographer	6	RTGS	credit	C761750706
18	saving	136350	Manager	6	Cheque	credit	C1237762639
19	saving	70500	CEO	1	Paypal	credit	C2033524545
20	saving	10000	Accountant	1	DD	credit	C1670993182
21	saving	13000	Bussiness	6	NEFT	credit	C20804602
22	saving	14000	Employee	7	Netbanking	credit	C1566511282
23	saving	60000	Professor	8	UPI	credit	C1959239586
24	saving	70000	IT	12	Googlepay	credit	C504336483
25	saving	30000	Software Engineer	13	Paytm	credit	C1984094095
26	saving	100000	HR	24	cash	credit	C1043358826
27	saving	45000	Doctor	12	RTGS	credit	C1671590089
28	saving	15000	Lawyer	23	Cheque	credit	C1053967012
29	saving	84000	Architect	67	Paypal	credit	C1632497828
30	saving	400000	Journalist	12	DD	credit	C764826684
31	saving	26000	Nurse	34	NEFT	credit	C2103763750
32	saving	83000	Carpenter	23	Netbanking	credit	C215078753
33	saving	87000	Politician	32	UPI	credit	C840514538
34	saving	62000	Photographer	43	Googlepay	credit	C1768242710
35	current	500000	Manager	2	Paytm	credit	C247113419
36	current	30000	CEO	3	cash	credit	C1238616099
37	current	150000	Accountant	4	RTGS	credit	C1608633989
38	current	20000	Bussiness	6	Cheque	credit	C923341586

Fig. 2. Raw data before subjected to Preprocessing

Fig 2, respresents the transaction data set created for this research work by accuring the knowledge form kaggle dataset and performing an survay on money laundering process from different bankes.

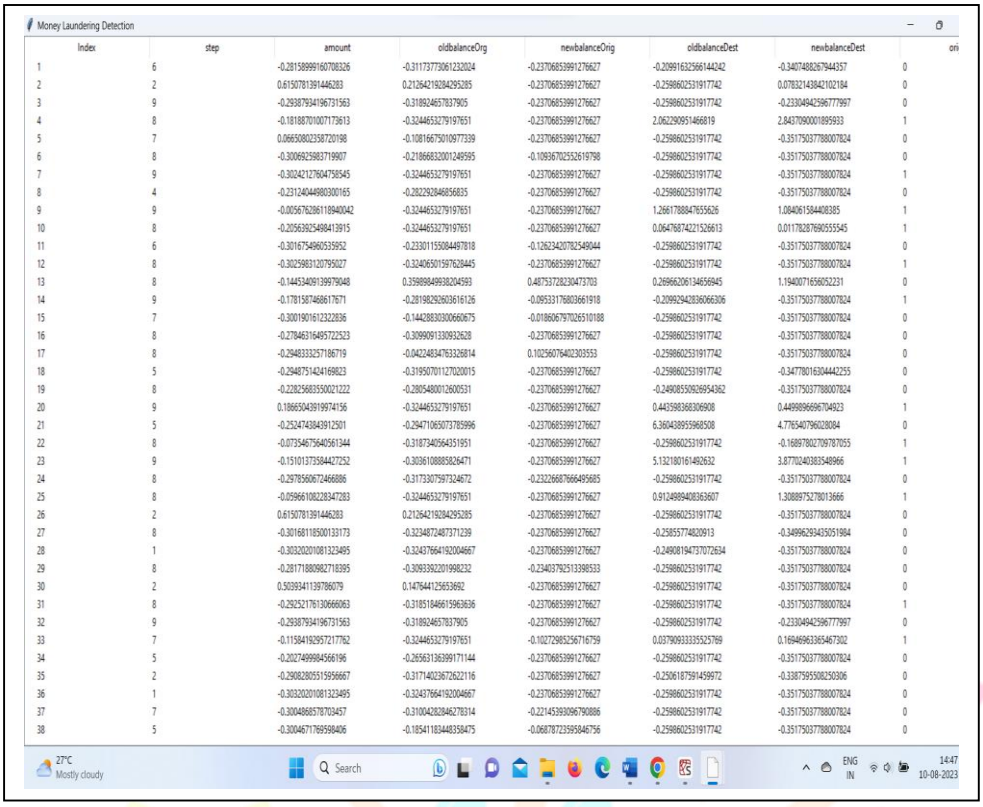


Fig. 3. Transaction data after Preprocessing

Fig 3, respresents the transaction data set obtained for this research work after performing preprocessing. For this research work several prepocessing techniques are applied, data cleaning, the StandardScaler, scalar.transform, keras.Tokenizer and keras.pad_sequences.

V. CONCLUSION

Making dirty money into clean money is the primary objective of money laundering. The part financial institutions play in money laundering cannot be overstated. The results presented in this research lend support to the detection of money laundering. Ie. The preprocessed data is furter subjected to classification process. which is performed by meachine learning algorithms like Decision Tree, Random Forest and Artificial Neural Network. Deep Learning with Explainable ANN can be used for Accurate detection of money laundering [1], is proposed but not implemented.

REFERENCES

[1] **[1]** D. V. Lindberg and H. K. H. Lee, "Optimization under constraints by applying an asymmetric entropy measure," *J. Comput. Graph. Statist.*, vol. 24, no. 2, pp. 379–393, Jun. 2015, doi: 10.1080/10618600.2014.901225.

[2] B. Rieder, *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam, Netherlands: Amsterdam Univ. Press, 2020.

[3] Dattatray Vishnu Kute, Biswajeet Pradhan, Nagesh Shukla and Abdullah Alamri, Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering-A Critical Review, IEEE 15 june 2021.

[4] Alnasser Mohammed, *Money laundering in selected emerging economies: Is there a role for banks?* Journal of Money Laundering Control, ahead-of-print(ahead-of- print), S. A. S. (2021) Vol. 24 No. 1, pp. 102-110.

[5] I. Boglaev, "A numerical method for solving nonlinear integro-differential equations of Fredholm type," *J. Comput. Math.*, vol. 34, no. 3, pp. 262–284, May 2016, doi: 10.4208/jcm.1512-m2015-0241.

[6] Al-Suwaidi, N. A., & Nobanee, *A survey of the existing literature and a future research agenda.* Journal of Money Laundering Control, Antimoney laundering and anti-terrorism fi nancing, 2020, Vol. 24 No. 2, pp. 396-426.

[7] Amara, I., Khlif, H., & El Ammari, Strength of auditing and reporting standards, corruption and money laundering: A cross-country investigation. Managerial Auditing Journal, 2020, Vol. 35 No. 9, pp. 1243-1259.

[8] F. A. T. Force. International Standards On Combating Money Laundering and TheFinancing Of Terrorism & Proliferation—FATF Recommendations, F.A.T.Force, Editor. 2020, Financial Action Task Force. 2021Global Money Laundering, crime, U.N.O.oD.a.., Vienna, Austria, Feb 2021.

[9] M. Tiwari, A. Gepp, and K. Kumar, A review of money laundering literature: The state of research in key areas, Vol. 32 No. 2, pp. 271-303, 2020.

[10] Ardizzi, G., De Franceschis, P., & Giammatteo,M. Cash payment anomalies and money laundering: An econometric analysis of Italian municipalities. International Review of Law and Economics, 2018.

[11] Ardizzi, G., Petraglia, C., Piacenza, M., Schneider, F., & Turati, G. (2012). *Estimating Money Laundering through a "Cash Deposit Demand"* approach. 1–27, 2012.

[12] Ardizzi, G., Petraglia, C., Piacenza, M., Schneider, F., & Turati, *Money Laundering as a crime in the fi nancial Sector: A new approach to quantitative assessment, with an application to Italy*. Journal of Money, Credit and Banking, 46(8), 1555–1590,2014.

[13] Arnone, M., & Borlini, L. International anti-money laundering programs: Empirical assessment and issues in criminal regulation. Journal of Money Laundering Control, 13(3),226–271, 2010.

[14] Ba, H., & Huynh, Money launderingrisk from emerging markets: The case of Vietnam. Journal of Money Laundering Control, 2018

[15] Badal-Valero, E., Alvarez-Jareño, J. A., & Pavía, J. M. *Combining Benford's Law and machine learning to detect money laundering*. An actual Spanish court case. Forensic Science International,2018.

[16] Canhoto, Leveraging machine learning in the global fi ght against money laundering and terrorism financing: An affordances perspective. Journal of Business Research, 1–12.2020.

[17] Chao, X., Kou, G., Peng, Y., & Alsaadi, F. E. (2019). *Behavior monitoring methods for trade based money laundering integrating macro and micro prudential regulation*, A case from China.Technological and Economic Development of Economy,25(6) 2019.9383

[18] Chen, Z., Van Khoa, L. D., Teoh, E. N.,Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning, 2018

[19] F. A. T. Force. International Standards On Combating Money Laundering and TheFinancing Of Terrorism & Proliferation—FATF Recommendations, F.A.T.Force, Editor. 2020, Financial Action Task Force. 2021Global Money Laundering, crime, U.N.O.oD.a.., Vienna, Austria, Feb 2021.

[20] Georgios Pavlidis, *Global sanctions against corruption and asset recovery: a European approach*, Journal of Money Laundering Control: Volume 26 Issue 1, at: issue 1 2nd march 2023.

[21] Mohammed Ahmad Naheem, Presenting a Legal and Regulatory Analysis of the United Arab Emiration'spast present and future legislation on combating Money Laundering and Terrorist Financing, article published on: 2 march 2023, 1368-5201

[22] Amara, I., & Khlif, Financial crime, corruption and tax evasion: A cross- country investigation. Journal of Money Laundering Control, 2018.